STAT 8020 R Lab 2: Multiple Linear Regression I

Put your name here

May 05, 2026

Contents

Housing Values in Suburbs of Boston

The Boston housing dataset was collected in 1978. Each of the 506 observations contains aggregated information about housing and neighborhood characteristics for suburbs of Boston, MA.

Data Source: Harrison, D. and Rubinfeld, D.L. (1978) Hedonic prices and the demand for clean air. **J. Environ. Economics and Management** 5, 81–102.

Load the dataset

Code:

```
library(MASS)
data(Boston)
```

For the purposes of this lab, we will use only the following variables for conducting data analysis:

1. `medv`: Median value of owner-occupied homes, in 1000s
2. `lstat`: Percentage of lower-status population
3. `rm`: Average number of rooms per dwelling
4. `crim`: Per capita crime rate by town

Code:

You can use the code below to extract these variables:

```
vars <- c("medv", "lstat", "rm", "crim")
data <- Boston[, vars]
```

Exploratory Data Analysis

Numerical summary

1. Use the `summary` command to produce various numerical summaries of each of the 4 variables under consideration.

Code:

Graphical summary

2. Make a boxplot for each variable

Code:

3. Briefly discuss the distribution of each variable based on the boxplots. Comment on center, spread, skewness, and potential outliers.

Answer:

4. Create a scatterplot matrix to explore the relationships among these variables

Code:

Model Fitting

Here we will use `medv` as the response and `lstat`, `rm`, `crim` as predictors.

Simple Linear Regression

5. Fit a simple linear regression model.

Among the predictors `lstat`, `rm`, and `crim`, select **one predictor** to build the *best* simple linear regression model for predicting `medv`. Compare the models using R^2 and the scatterplots. Select the predictor that gives the best simple linear regression model and justify your choice.

Code:

6. Write down the fitted linear regression equation.

Answer:

Multiple Linear Regression

7. Fit a multiple linear regression using all predictors.

Code:

8. Write down the fitted linear regression equation

Answer:

9. Perform an overall F-test for the multiple linear regression model.

State: * Null and alternative hypotheses

- Test statistic
- p-value
- Decision
- Conclusion in context

Code:

Answer: